Self-serving, altruistic and spiteful lying in the schoolyard *

Isabelle Brocas
University of Southern California and CEPR

Juan D. Carrillo
University of Southern California and CEPR

April 2021

Abstract

Subjects 5 to 17 years old participate in two rounds of a lying game. In each round, each participant reports one outcome for themselves and one outcome for another random, anonymous person of the same age. We observe frequent but not omnipresent over-reports for oneself (self-serving lying). We also observe small aggregate levels of under-reports for others and a strong positive correlation across rounds in the reports for others, which might be explained by a coexistence of altruistic liars and spiteful liars in our population. Behavior is similar across ages, except for middle schoolers who exhibit a slightly higher inclination towards self-serving and spiteful lying. A focused analysis of choice in middle school reveals some differences by academic performance and socioeconomic status.

Keywords: developmental decision-making, field experiment, lying, altruism, spitefulness.

JEL Classification: C93.

*We are grateful to members of the Los Angeles Behavioral Economics Laboratory (LABEL) and three anonymous referees for their insights and comments and to Chris Crabbe for his excellent programming skills. We thank Emmanuelle Acker, Nordine Bouriche, Anneli Harvey and Mathieu Mondange from LILA and Adriana Díaz from KING for their help and support in running the experiments in their school. The study was conducted with the University of Southern California IRB approval UP-12-00528. We acknowledge the financial support of the National Science Foundation grant SES-1851915. Address for correspondence: Juan D. Carrillo, Department of Economics, University of Southern California, 3620 S. Vermont Ave., Los Angeles, CA 90089, USA, <juandc@usc.edu>.

1 Introduction

Lying and other-regarding preferences are two paradigms that have received much attention in the recent experimental literature. The research on lying has emphasized, among other things, that lying is frequent but by no means universal. It depends on whether outcomes are observed. Also, many individuals provide partial lies, suggesting the existence of a direct, a social and a reputational cost of lying (see Abeler et al. (2019) for a review). The other-regarding literature shows that total and relative outcomes matter but also that intentions and perceptions are important. It highlights that most individuals satisfy basic rationality in their concern for others, and that outcomes in social games depend on the set of rewards and punishments available (see Cooper and Kagel (2016) for a review). It has been also emphasized that children evolve with age in their degree of altruism, generosity and spitefulness (Fehr et al., 2008, 2013; Cobo-Reyes et al., 2019).

Research in psychology (Gingo et al., 2017) has demonstrated different propensities to lie depending on the action domain: moral issues (e.g., related to the rights and welfare of others), social-conventional issues (e.g., societal norms and behavioral uniformities), and prudential issues (e.g., concerns with safety and comfort). Lying is an increasing concern for parents, teachers and policy makers. It is a critical problem since numerous studies have shown that this personality trait is persistent (Stouthamer-Loeber, 1986; Gervais et al., 2000). It is also associated with other adverse behaviors such as aggressiveness and destructiveness (Stewart and de Blois, 1985; Gervais et al., 1998) and with negative emotions such as depression (Dykstra et al., 2020). Perhaps more importantly, lying is a predictor of future delinquency (Mitchell and Rosa, 1981).

And yet, not all types of lying or cheating are perceived equally by society. Cheating for one's benefit (e.g., plagiarizing an essay) is different from cheating for someone else's benefit (e.g., anonymously posting one's essay for others to use) which is itself not the same as cheating for someone else's detriment (e.g., the same posting but with deliberate errors and omissions). In this paper, we combine the literatures on lying and social preferences in a population of children and adolescents to address the following questions. How likely are individuals to lie in decisions that affect only other people? Will they lie for their benefit or for their detriment? Does the amount and type of lying change with age? Is there a link between lying for oneself and lying for others?

To address these questions, we adopt Fischbacher and Föllmi-Heusi (2013)'s elegant paradigm, where subjects privately roll a dice, report the outcome and are paid as a function of the report (from 1 to 6). We ask a population of children and adolescents from a private school in Los Angeles to play two consecutive rounds with two dice, a red and a green. In each round, the points in the red dice translate into a payoff for themselves and the points in the green dice translate into a payoff for another anonymous person.

The experiment yields the following conclusions. First, and as in most of the existing literature (Bucciol and Piovesan, 2011; Glätzle-Rützler and Lergetporer, 2015; Maggian and Villeval, 2016; Amato et al., 2019; Alan et al., 2019; Zhang et al., 2019), children lie often – but not always – for themselves. Lying propensities in our experiment are similar at all ages; only middle schoolers engage in slightly more self-serving lying than the other age groups. Surprisingly, reports for themselves are uncorrelated across rounds, which is at odds with standard theories of constrained optimal choice given a fixed (direct or reputational) cost of lying. Second, and regarding the report for others, we observe at the aggregate level a small amount of under-reporting and a similar behavior at all ages except for middle schoolers who, once again, lie slightly more. In contrast with reports for oneself, reports for others are positively and strongly correlated across rounds. This result could be explained by individual heterogeneity, with a coexistence of a small fraction of altruistic liars (who systematically over-report the outcome destined to others) and a somewhat larger fraction of spiteful liars (who systematically under-report the outcome destined to others). Finally, we ran the same experiment with a larger sample of middle schoolers from a public school in the same neighborhood. We differentiate between two tracks, publicly categorized by the school: "challenged" and "regular". Challenged students are children with behavioral or learning difficulties. These students are predominantly from low SES households. Regular students are in either standard or honors classes, and they belong to higher SES households. We found some differences between these two tracks. Challenged students lie very significantly to their own benefit and do not lie for or against others, whereas participants in the regular track provide lower reports in both cases, that is, they lie (but less) for themselves and they lie negatively for others.

Naturally, we are not the first to study lying for others and lying by children. On social lying, there is a strand of research initiated with Erat and Gneezy (2012) that studies strategic information transmission in sender-receiver games when the payoffs of both players are affected by the action of the latter. The literature shows that some subjects are averse to lying even if it improves the payoff of both subjects, while others are willing to lie to benefit others even if it comes at a (small) cost for themselves. The subsequent research has demonstrated in-group favoritism in lying (Cadsby et al., 2016), as well as increased lying when it benefits both oneself and others (Gino et al., 2013).

On developmental lying, there is also a recent literature which shows that children 5 to

15 years old lie significantly and similarly at all ages (Bucciol and Piovesan, 2011). Socially oriented (Maggian and Villeval, 2016) and inequality averse (Amato et al., 2019) children lie less than selfish or envious ones. Among elementary school children, participants with higher cognitive ability and higher socioeconomic status also lie more frequently (Alan et al., 2019), and in a children variant of Erat and Gneezy (2012), Glätzle-Rützler and Lergetporer (2015) show that 5^{th} graders lie more than 11^{th} graders but only when they improve the payoffs of both subjects. Finally, Zhang et al. (2019) also study selfish and social lying in children and adolescents. The paper, however, focuses on the effect of priming (self-identification as a 'good' or a 'bad' guy) and the impact of peer behavior (the overheard reports of previous children in the group).[1] To our knowledge, no study in the adult or children literature has analyzed the issue of purely altruistic vs. purely spiteful lying, that is, lying to benefit vs. lying to hurt others in a basic, non-signaling game where the choice for others is stripped of consequences for oneself. While many everyday decisions to lie have some strategic component (e.g., letting others copy out the homework may increase status or foster reciprocity) our design provides a pure measure of the intrinsic tendency to lie absent those considerations. Finally, but importantly, our setting also provides a benchmark to investigate whether a person's inclination to lie is intrinsic and independent of the object of the lie or, instead, if such motivations are context dependent. Indeed, the literature usually assumes that the act of lying entails a personal moral costs. By studying the correlation between lying for oneself and lying for others, we can determine whether choices stem from similar trade-offs –and are therefore motivated by the same underlying moral principles– or whether they are situationally-dependent.

2 Design and theory

We conduct a social variant of the lying game introduced in the seminal work of Fischbacher and Föllmi-Heusi (2013), in a population of children and adolescents. We implement a modified set of procedures compared to standard economics laboratory experiments run with adults. A review of the methodological challenges that arise with a population of children together with some general guidelines to address those obstacles can be found in Brocas and Carrillo (2020).[2] We closely follow those general principles in this paper.

[1] See also List et al. (2018) and Sutter et al. (2019) for recent surveys of economic experiments with children.

[2] In a nutshell, the principles are: (i) adapt the length and procedures to a population with limited attention span; (ii) offer age-appropriate incentives (possibly different at different ages); (iii) present the task in a way that subjects are not required to possess strong analytical skills to participate (e.g., graphical

2.1 Experimental design and procedures

Participants. We recruited 354 participants from K to 11$^{\text{th}}$ grade from Lycée International de Los Angeles (LILA), a French-English bilingual private school in Los Angeles. We also recruited 211 middle schoolers (grades 6, 7 and 8) from Thomas Starr King Middle School (KING), a public school in Los Angeles. For comparison, we ran the experiment with a control population of 60 USC college undergraduates (U) using the same procedures. We present a discussion of the differences across populations in the analysis section. We lost data from 3 participants at LILA due to a problem with a computer. Table 1 reports the remaining 622 participants by grade.

	LILA												KING			USC
Age range	5-8 (**G1**)			8-11 (**G2**)			11-14 (**G3**)			14-17 (**G4**)			11-14			18+
Grade	K	1	2	3	4	5	6	7	8	9	10	11	6	7	8	U
# subjects	38	37	43	34	17	20	41	27	18	24	23	29	122	46	43	60

Table 1: Summary of participants by grade and location.

Tasks. The experiment consisted of two tasks always performed in the same order. We started with the "dice game" studied in this paper. After a short break, we performed 18 rounds of a three-person simultaneous play, dominance-solvable game. We report and analyze the results of the second game in a different paper (Brocas and Carrillo, 2019b). Subjects did not learn the points accumulated in either task until the end of the experiment. Naturally, they knew their own reports in the dice game.

General procedures. We ran 30 sessions at LILA, 19 sessions at KING and 5 sessions at USC, each with 9 to 15 subjects. All the interaction between subjects was computerized through touchscreen PC tablets. Tasks were programmed with the open source software 'Multistage Games.'[3] Sessions with school-age students were run in a classroom at the school whereas sessions with undergraduates were run at the Los Angeles Behavioral Economics Laboratory (LABEL) at USC. Procedures were *identical* in all cases, except for payments as explained below. For each school-age session, we tried to have male and female subjects from the same grade, but for logistic reasons we sometimes mixed subjects of two consecutive grades.

The dice game. Given the very young age of some of our participants, misunderstanding

interfaces and simple instructions); (iv) understand, describe and compare the children population, and (v) include a benchmark adult comparison group whenever possible.

[3] Downloading instruction can be found at `http://ssel.caltech.edu:8000/multistage`.

the fundamentals of the game is a real possibility due to cognitive limitations or inattention (Brocas and Carrillo, 2020). We therefore put very strong emphasis on presenting the game using a simple, accessible, non-analytical framework that children as young as 5 years old could easily grasp. To this purpose, we gave participants a physical cup with two dice, one red and one green. We instructed them to shake the cup, privately overturn it, and record in a touchscreen computer the numbers displayed on their dice. The number on the red dice corresponded to points for themselves (from now on, the "ME" choice). The number on the green dice corresponded to points for another person randomly and anonymously drawn from all participants in the session (from now on, the "OTHER" choice). After all participants had finished (and without telling them the points they received from the other participant), the experimenter announced a second and final round of the same game with new, random and anonymous partners.[4] During the experiment, we made sure that decisions were private. Subjects secretly shook their dice in the cup and had cardboard separators between stations. The experimenter(s) remained in a corner of the room during the task. Also, the screen became blank once the subject had input the decision and pressed OK. We emphasized that it was not an exchange game: the person from whom the subject received points was random and anonymous (in our words, "selected by the computer without telling it to anyone") and *always different* from the person to whom the subject conferred points.[5] Figure 1 presents a screenshot of the game. We report a transcript of the instructions in Appendix B.[6]

Payoffs. During the experiment, subjects accumulated points. Incentives were calibrated to account for differences in marginal value of money and opportunity cost of time across ages rather than to equalize the payment scheme (see Brocas and Carrillo (2020) for a discussion). We implemented three different payment rules depending on the population.

[4]As extensively discussed in the literature (Gneezy et al., 2018; Abeler et al., 2019), privacy (non-observed game as opposed to observed game) reduces demand effects but prevents the analysis of lying at the individual level. Playing two rounds increases the statistical power of the group analysis. It allows us to detect persistence of behavior and to test the hypothesis of an intrinsic individual cost of lying.

[5]We considered the possibility of having only one-half of the children play the dice game to completely eliminate any reciprocity consideration but found it impractical due to the dramatic reduction in sample size. We also considered the option of having recipients from a different school but decided against it because payments would look distant and uncertain. More importantly, we noticed during pilot testing that an advantage with young children of playing both roles is that it makes the choice more concrete, that is, they can better relate to what it means being a recipient in this game.

[6]Given the simplicity of the setting and to avoid framing or anchoring effects, we decided not to include a comprehension quiz. Instead, we had a brief, oral interactive conversation where we asked participants if they understood the game and had any question or concern. The only issue raised in some sessions was a clarification that one's recipient of points was never the same person as one's conferrer of points.

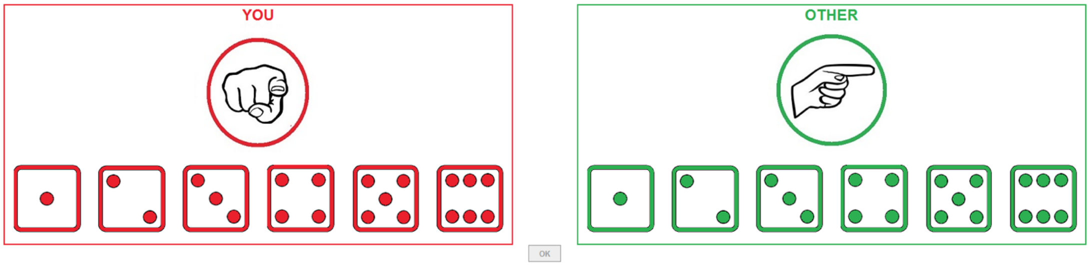

Figure 1: Screenshot of the *Dice Game*. Participants tapped on the numbers corresponding to the red dice (points for themselves, graphically represented with a finger pointing inwards) and the green dice (points for another participant, graphically represented with a finger pointing sideways). After they pressed "OK", the screen was replaced by a "please wait" sign.

USC students, and school-age subjects from grade 6 and above had points converted into money paid immediately at the end of the experiment in cash (USC) and with an amazon e-giftcard (LILA and KING, where cash transfers on premises are not allowed). For the dice game studied in this paper, USC subjects accumulated $0.40 per point reported in the dice (either to themselves or from the other participant). LILA and KING subjects accumulated $0.30 per point. This means a range of payments of $1.60 to $9.60 (USC) and $1.20 to $7.20 (LILA and KING). Subjects also accumulated points for the other task.[7] The dice game lasted 10 minutes and the entire experiment never exceeded one school period (50 minutes), including instructions and payments.

For elementary school subjects (grades K to 5), we set up a shop with 20 to 25 pre-screened, age appropriate toys and stationary.[8] Different toys had different point prices. Before the experiment, children were taken to the shop and showed the toys they were playing for. They were also instructed about the point price of each toy and, for the youngest subjects, we explicitly stated that more points would result in more toys. At the end of the experiment, subjects learned their point earnings in the two games and were accompanied to the shop to exchange points for toys. We made sure that every child earned enough points to obtain at least three toys, although there was significant variance in the number and value of toys selected.[9]

[7] There was also a $5 show-up fee paid only to USC students. Average earnings for the whole experiment (not including show-up fees) were $15.2 (USC), $11.2 (LILA) and $10.8 (KING).

[8] These included gel pens, bracelets, erasers, figurines, die-cast cars and trading cards for younger kids, and apps, calculators and earbuds for older kids. Children, however, could choose any item they wanted and some of them, like the fidget spinner, were popular for boys and girls of all ages.

[9] The procedure emphasizes the importance of accumulating points while making the experience enjoyable. At this age, a toy is also a significantly more attractive reward than money. Children are familiar

Other information. At the end of the experiment, we also collected demographic information consisting of "gender", "age", "grade", "number of siblings" and "favorite subject at school".

2.2 Theoretical considerations

The recent literature has proposed numerous theoretical models that incorporate direct, social and reputational costs of lying into the traditional selfish, profit maximizing framework (see e.g., Gneezy et al. (2018); Dufwenberg and Dufwenberg (2018); Khalmetski and Sliwka (2019); Brocas and Carrillo (2019a) as well as Abeler et al. (2019) which encompasses and extends many existing theories). These sophisticated models typically consider one-shot decisions and lying for oneself, so they do not directly apply to our framework. While the purpose of the present work is not to provide a new theory to address our extended setting, some considerations are in order.

First, the most common (though certainly not only) way to explain behavioral heterogeneity in this literature is to assume that the cost of lying differs across individuals. We will implicitly use this assumption in our discussion and interpretation of results.

Second, a model where each individual (i) has a personal cost of lying which is the same (or, at least, positively correlated) over time and (ii) has stable preferences (therefore incentives to lie in the same direction in both rounds) immediately implies positive correlation of ME choices across rounds and positive correlation of OTHER choices across rounds. Indeed, while there are several reasons why incentives to lie may not be identical across rounds (stochasticity, convex accumulation cost, etc.), standard theories share the conclusion that individuals with low costs are more likely to lie in both rounds and individuals with high costs are more likely to lie in no round. Interestingly, however, lying may take different forms. Following the existing theoretical models and empirical findings, we will assume that individuals will only lie in ME to improve their payoffs, thereby tilting the distribution of reports in each round towards high values. By contrast, given the documented coexistence of individuals with altruistic and spiteful preferences especially in children (Fehr et al., 2008; Brocas et al., 2019), non-truthful reports in OTHER can take the form of both positive / altruistic lying (over-reports) and negative / spiteful lying (under-reports). In that case, aggregate departures may not be apparent whenever there are individual departures in both directions. Our two-round setup provides an extra tool to detect lying in this case, namely a positive correlation across rounds.

---

with this method of accumulating points (or tickets) that are subsequently exchanged for rewards since it is commonly employed in fairs and arcade rooms. We spent an average of $4 in toys per child.

Third, if lying has a common cost component for the individual which is independent of the object of the lie (ME or OTHER), then we should observe a positive correlation of individual lying across choices. However, and as discussed above, detecting such correlation may not be easy when different individuals lie in different directions. A prediction is that we should observe a correlation between high reports in ME and extreme (high and low) reports in OTHER.

3 Selfish and social lying: from kindergarten to adulthood

We first focus on the LILA population (grades K to 11) and treat USC as the adult control group. Many economic experiments with children do not perform the same experiment with an adult population. They rely on previous research for the adult reference or refrain from making comparisons between children and adults. We believe it is helpful to include an adult control group using identical procedures to establish a behavioral benchmark (Brocas and Carrillo, 2020). At the same time, it is key to recognize its limitations. In our case, the majority of students at LILA are Americans and Europeans from caucasian families of upper-middle socioeconomic status. After high school, they typically go to well-ranked colleges in Europe and North America (including USC). It is therefore an imperfect but reasonable match for the USC population, despite some differences in individual characteristics (nationality, family background, size of peer group, etc.).

For statistical power, we group our LILA subjects in four naturally clustered age groups (see Table 1): younger elementary **G1** (K-1st-2nd; 118 subjects from 5 to 8 years old), older elementary **G2** (3rd-4th-5th; 71 subjects from 8 to 11 years old), middle school **G3** (6th-7th-8th; 86 subjects from 11 to 14 years old), and high school **G4** (9th-10th-11th; 76 subjects from 14 to 17 years old). The final group is our undergraduate comparison population **U** (60 subjects).

It is important to note that the robustness of results in a non-observed game is compromised if samples are relatively small. It is therefore important to provide a power analysis. Assume truthful reports as the null hypothesis. This translates into a 1/6 probability of obtaining each outcome in each round. To mirror the analysis that follows, we consider χ^2-tests in which total reports are grouped in 3 categories (2-4, 5-9 and 10-12). Assuming a significance level of 0.05 and power of 0.80, we look for the sample size needed to detect different effect sizes (Cohen, 2013). Effect sizes are heuristically categorized as small (0.1), medium (0.3) or large (0.5). In order to detect effect sizes of 0.5, 0.3 and 0.1, the sample size needed is $n = 963$, $n = 107$ and $n = 38$, respectively. With a sample size between

70 and 110 per school-age group, target effect sizes are between 0.30 and 0.37. Further details of the power analysis can be found in Appendix A1.

3.1 Aggregate behavior: selfish and social lying

Table 2 summarizes the average number of points in each age group that participants allocate to themselves (ME) and to the other participant (OTHER). We report the points both by round (1 and 2), as well as combining both rounds together (all). The last column reports the theoretical prediction under truthful reports.

	G1	**G2**	**G3**	**G4**	**G-All**	**U**	Theory
ME-1	4.02 (1.64)	4.14 (1.75)	4.35 (1.45)	4.29 (1.70)	4.18 (1.63)	4.18 (1.76)	3.5
ME-2	4.31 (1.66)	4.56 (1.67)	4.47 (1.62)	4.41 (1.66)	4.41 (1.65)	4.23 (1.71)	3.5
ME-all	8.32 (2.50)	8.70 (2.39)	8.81 (2.20)	8.70 (2.48)	8.60 (2.40)	8.42 (2.68)	7.0
OTHER-1	3.54 (1.73)	3.46 (1.76)	3.22 (1.62)	3.53 (1.77)	3.44 (1.71)	3.60 (1.85)	3.5
OTHER-2	3.58 (1.73)	3.66 (1.74)	3.33 (1.61)	3.51 (1.76)	3.52 (1.71)	3.22 (1.67)	3.5
OTHER-all	7.13 (2.70)	7.13 (2.52)	6.55 (2.50)	7.04 (2.71)	6.97 (2.67)	6.82 (2.56)	7.0

(standard deviations in parenthesis)

Table 2: Average choices for ME and OTHER by round and grade

From these simple aggregate statistics, we can already notice significant lying in ME in all age groups with a slight hump-shaped behavior with age. Conversely, there is no evidence of aggregate lying in OTHER, except for **G3** who have a slight tendency to under-report. The distribution of aggregate behavior is not statistically different between rounds 1 and 2 in any age group. This is true both for ME and for OTHER (Wilcoxon Signed Rank test, all $p > 0.05$). Given the similarity of behavior across rounds, for the analysis below we will consider the sum of points in the two rounds, which ranges from 2 to 12.

Figure 2 reports the total allocation to ME (left) and OTHER (right) by age group. Since we know from previous research that lies do not always take the form of extreme reports (e.g, a dishonest report of 5 rather than 6 is not uncommon) but we also want to provide a simple visual representation, we divide the reports into three categories: 2 to 4 points, 5 to 9 points and 10 to 12 points. The theoretical proportions for these categories are 0.167, 0.667 and 0.167, and represented by the dashed lines in Figure 2.[10]

Behavior across age-groups. With the exception of **G3**, participants behave similarly

[10]The choice is obviously ad-hoc. However, we figured that the fraction of lying vs. truthful report for individuals who report, say, 5+4 is likely to be low. In Appendix A2 we provide the entire distribution of reports. We find no notable differences but, naturally, a much weaker statistical power.

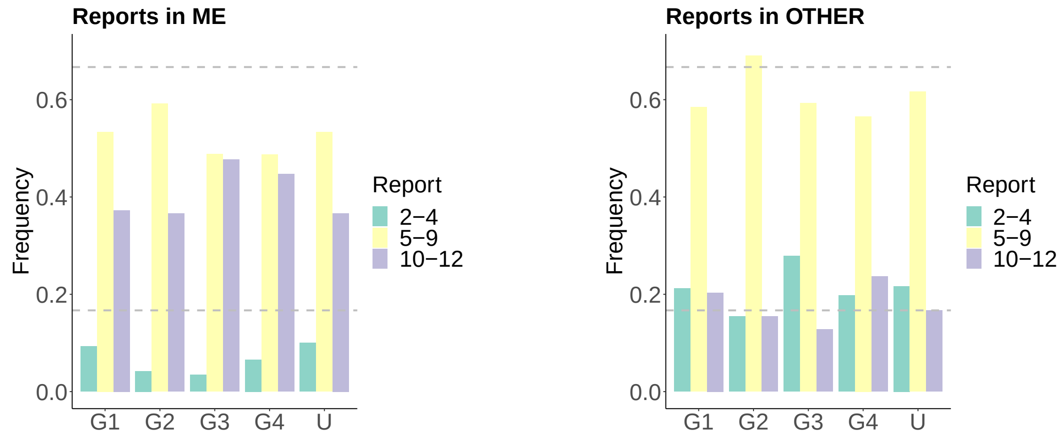

Figure 2: Frequency of allocations by age group to ME (left) and to OTHER (right)

across all school age-groups in ME and across all school age-groups in OTHER (pairwise χ^2-tests, all $p > 0.05$). They also behave similarly to the adult control group in both cases (χ^2-tests, all $p > 0.05$). Participants in **G3** behave significantly different from **U** in ME (χ^2-test, $p = 0.011$). The difference in behavior between **G3** and **G1** in ME is close to statistically significant (χ^2-test, $p = 0.051$). **G3** also behave differently from **G2** and **G4** in OTHER (χ^2-tests, $p = 0.006$ and $p = 0.011$), by submitting significantly more low reports.[11] The similarity of behavior across age groups in ME is reminiscent of the original article by Bucciol and Piovesan (2011). We also share with Zhang et al. (2019) the result that lying peaks in middle school.[12] On the other hand and in contrast to our findings, Glätzle-Rützler and Lergetporer (2015) and Maggian and Villeval (2016) find that 4^{th}-5^{th} graders lie more that 11^{th} graders and 7^{th}-8^{th} graders respectively. Overall, the literature does not seem to show a fully consistent pattern regarding the evolution of lying from childhood to adolescence, perhaps because differences are only marginal and depend on the paradigm and procedure.

Behavior in ME. There is very significant self-serving lying in all age groups: 2-4 and 5-9 are under-represented whereas 10-12 is very strongly over-represented (χ^2-tests comparisons with theoretical probabilities, $p < 0.001$). At the same time, lying is not

[11] χ^2-tests are used to compare the proportions in each category of an age-group with the proportions in each category of another age group. Multinomial exact tests are used as a robustness check and confirm the results.

[12] Notice that **G3** is the youngest group paid in cash. Their different choices may be due in part to higher marginal incentives.

omnipresent: in all age groups, more than 50% of participants report values between 2 and 9 despite the absence of monitoring and penalty. This result is qualitatively similar to those found in the literature both with children and with adults.

Behavior in OTHER. The overall distribution is not statistically different from the theoretical triangular one in any age group with the exception of **G3**. In that age-group, 2-4 is over-represented (28.0%) at the expense of 5-9 and 10-12 (χ^2-test comparison with theoretical probabilities, p = 0.018). When we pool together all school-age groups, there is a small but statistically significant tendency towards negative lying: 2-4 is over-represented (21.4%) whereas 5-9 is under-represented (60.4%) and only 10-12 is close to the theoretical proportion (18.2%) (χ^2-test comparison with theoretical probabilities, p = 0.027).[13]

As discussed in section 2.2, in a non-observed game, small (or no) aggregate deviations from predictions under truthful revelation in the choice for OTHER does not necessarily imply small (or no) lying. The analysis in the next sections aims at separating different types of individual behavior.

Given the similarity across age groups, for the remaining of the section, we pool together the 351 school-age participants (K through 11) and remove from the sample the control adult group. This overlooks the documented differences of **G3** compared to the other groups. To address the singularity of the middle school population, we provide in section 4 a focused analysis of lying in this age group, and compare it to children of the same age in another school.

3.2 Individual behavior across choices

To investigate the relationship at the individual level between choice in ME and choice in OTHER, we present in Table 3a the proportion of individuals whose total report is 2-4, 5-9 and 10-12 in ME (rows) and 2-4, 5-9 and 10-12 in OTHER (columns), with the aggregate proportions of ME and OTHER reported in the last column and the first row, respectively. For comparison, we present in Table 3b the proportions if every participant reported truthfully.

Consistent with Figure 2, there is very significant over-reporting in ME: 41.3% of high reports (10-12) instead of the expected 16.7% at the expense of both intermediate (5-9) and low (2-4) reports. There is also a slight over-tendency towards low reports in OTHER (21.4% of 2-4). The excessive tendency to provide high reports in ME is equally frequent for individuals with intermediate reports in OTHER (41.5%) than for individuals with extreme

[13] Results are similar if we consider the entire distribution of reports instead of clustering it in three categories (Kolmogorov-Smirnov test, p = 0.001).

		0.214	0.604	0.182	
	10-12	0.085	0.251	0.077	0.413
ME	5-9	0.117	0.311	0.097	0.524
	2-4	0.011	0.043	0.009	0.063
		2-4	5-9	10-12	
			OTHER		

(a) Empirical reports

		0.167	0.667	0.167	
	10-12	0.028	0.111	0.028	0.167
ME	5-9	0.111	0.444	0.111	0.667
	2-4	0.028	0.111	0.028	0.167
		2-4	5-9	10-12	
			OTHER		

(b) Theoretical reports

Table 3: Reports across choices: ME (rows) and OTHER (columns).

reports (2-4 and 10-12) in OTHER (41.0%) (test of comparison of proportions, $p = 1.0$). Similarly, the slight tendency to provide low reports in OTHER is also equally frequent for individuals with high reports in ME (20.7%) than for individuals with intermediate and low reports in ME (21.8%) (test of comparison of proportions, $p = 0.899$).[14] Overall, and contrary to our theoretical prediction, there is no evidence of a correlation between lying in ME and lying in OTHER. This suggests that if individuals suffer a cost of lying as the theoretical literature often posits, this cost is not universal and, instead, depends crucially on the context of the lie.[15] In our experiment, the individual cost of lying for oneself appears to be independent of the cost of lying for someone else, a plausible but somewhat surprising conclusion. Notice also that the absence of correlation is consistent with individuals with no inequality concerns. More likely, however, it may mask a combination of inequality-prone and inequality-averse individuals.[16]

3.3 Individual behavior across rounds

Next, we look at the behavior across rounds separately for each choice. We denote by R_t^c the report R in round $t \in \{1, 2\}$ of choice c, with $c = m$ (ME) or $c = o$ (OTHER). We group the reports into three categories: $R = L$ (low, 1-2), $R = M$ (medium, 3-4) and

[14] The fraction of high reports in OTHER is also equally frequent for individuals with high reports in ME (18.6%) than for individuals with intermediate and low reports in ME (18.0%) (test of comparison of proportions, $p = 0.986$).

[15] The result relates to Biziou-van Pol et al. (2015) who find that dictator donations are positively correlated with altruistic lies (hence, driven by altruism) but negatively correlated with Pareto white lies (hence, driven by selfishness).

[16] While social preferences and preferences for truth-telling are likely linked, studying their relationship would require developing a new theoretical framework.

$R = H$ (high, 5-6). Table 4a presents the proportion of school-age participants who report (R_1^m, R_2^m) in ME. It also reports the total proportions in round 1 (last column) and round 2 (first row). If reports are truthful, all combinations should be equally likely within rounds $(\Pr(R_1^m, R_2^m) = 1/9\ (\simeq 0.111))$. We use this information to report $\Pr(R_2^c \mid R_1^c)$ in Table 4b, that is, the conditional probabilities of choice in round 2 as a function of choice in round 1. If reports are uncorrelated across rounds, these conditional probabilities should be equal $(\Pr(R_2^m \mid L_1^m) = \Pr(R_2^m \mid M_1^m) = \Pr(R_2^m \mid H_1^m))$.

		0.162	0.254	0.584	
	H_1^m	0.071	0.131	0.291	0.493
1	M_1^m	0.057	0.074	0.194	0.325
	L_1^m	0.034	0.048	0.100	0.182
		L_2^m	M_2^m	H_2^m	
			2		

(a) Joint probabilities $(\Pr(R_1^m, R_2^m))$

	$\Pr(L_2^m \mid \cdot)$	$\Pr(M_2^m \mid \cdot)$	$\Pr(H_2^m \mid \cdot)$
H_1^m	0.144	0.266	0.590
M_1^m	0.175	0.228	0.597
L_1^m	0.187	0.264	0.549

(b) Conditional probabilities $(\Pr(R_2^m \mid R_1^m))$

Table 4: Reports across rounds in ME.

The choices in ME represented in Table 4a constitute a behavioral template of heterogeneous self-serving lying. The combinations yielding higher expected gains are associated with more frequent reports. The range is widespread from most over-reported $(\Pr(H_1^m, H_2^m) = 0.291)$ to most under-reported $(\Pr(L_1^m, L_2^m) = 0.034)$. There are more H-reports in the second round than in the first round (0.584 vs. 0.493, two-sample test for equality of proportion, p = 0.019). However, as we can see from Table 4b, the distribution of H-reports in round 2 does not depend on the reports in round 1: $\Pr(H_2^m \mid H_1^m) = 0.590$, $\Pr(H_2^m \mid M_1^m) = 0.597$ and $\Pr(H_2^m \mid L_1^m) = 0.549$ (three-sample test for equality of proportion, p = 0.80). There are no differences in the conditional distributions of M- and L-reports in round 2 either. Overall, the empirical correlation of ME choices across rounds is small and not statistically significant (PCC = 0.07, p = 0.183), which is highly surprising.[17] As discussed in section 2.2, a standard model with an individually determined cost of lying and stable preferences would result in positive correlation of high reports across periods: an individual who over-reports in round 1, due in part to a low cost of lying,

[17] If we do not include our youngest participants (**G1**), the correlation increases (PCC = 0.09, p = 0.067), although it remains low and not significant at conventional levels.

should be prone to also over-report after learning the existence of round 2.[18]

This documented lack of correlation may be simply due to the limited dynamic data available in our study since, after all, we only consider two rounds. Indeed, some participants may attempt to "mask" a high report in the first round with a lower report in the second round. Some others may realize only after the first round that actions are private. These participants may be prompted to switch from truthful first round reports to second round lies.[19] Such conflicting considerations would not be strong enough to mask a positive correlation if individuals played enough rounds but it might suffice to hide it in our two-round setup.

We next present in Tables 5a and 5b the same information across rounds regarding the choice in OTHER, (R_1^o, R_2^o).

		0.334	0.316	0.350	
	H_1^o	0.094	0.085	0.125	0.304
1	M_1^o	0.100	0.111	0.134	0.345
	L_1^o	0.140	0.120	0.091	0.351
		L_2^o	M_2^o	H_2^o	
			2		

(a) Joint probabilities $(\Pr(R_1^o, R_2^o))$

	$\Pr(L_2^o \mid \cdot)$	$\Pr(M_2^o \mid \cdot)$	$\Pr(H_2^o \mid \cdot)$
H_1^o	0.309	0.280	0.411
M_1^o	0.290	0.322	0.388
L_1^o	0.399	0.342	0.259

(b) Conditional probabilities $(\Pr(R_2^o \mid R_1^o))$

Table 5: Reports across rounds in OTHER.

The occurrence of each report combination in OTHER (Table 5a) is not statistically different from what we would observe under truthful revelation (0.111) (multinomial test of comparison with theoretical distribution, $p = 0.284$). This is consistent with the (small) aggregate level of lying in OTHER choices. Aggregate behavior is also very similar in rounds 1 and 2. However, in support of our theoretical prediction, choices in OTHER are positively and very significantly correlated across rounds (PCC = 0.17, p = 0.001). The correlation is particularly important when we study individuals with low reports in round 1: $\Pr(L_2^o \mid L_1^o) = 0.399$ vs. $\Pr(H_2^o \mid L_1^o) = 0.259$ (test of comparison of proportions,

[18] The effect would be increased if the marginal cost of a second lie is lower than that of a first lie (as, for example, in Tirole (1996)) and decreased (but not removed) if the cost of lying depends on the "stock" of lies accumulated (as, for example, in Brocas et al. (2021)).

[19] In Appendix A3 we show that individuals with an L_1^m report do not exhibit lying tendencies in OTHER choices.

$p = 0.030$). It suggests that our participants deviate from truthful reports when it concerns other people. However, since deviations take opposite forms for different individuals (altruistic vs. spiteful), they do not show at the aggregate level as strikingly as in the case of ME choices. While the literature (Erat and Gneezy, 2012; Glätzle-Rützler and Lergetporer, 2015; Maggian and Villeval, 2016) has already emphasized individual propensities to engage in different types of lies (white, black, altruistic, Pareto improving), our paper differentiates in one single setting between subjects who lie to benefit others and those who lie to hurt others when the choices have no direct consequences for themselves.[20] At the same time, the result is not entirely surprising given the documented coexistence of children with altruistic and spiteful social preferences (Fehr et al., 2008, 2013; Brocas et al., 2019).

3.4 Regression analysis

We next perform some regressions to further investigate the decisions of individuals across choices and rounds. In Table 6, we consider all school-age participants and present OLS regressions at the individual level of ME choice in the second round as a function of ME and OTHER choices in the first round (columns 1, 2 and 3) as well as ME choices in both rounds (column 4). We include as individual control variables the gender, whether the participant has siblings and the age (either in months or as dummy age categories with **G1** as the omitted variable). We also present the same information regarding OTHER choices (columns 5 to 8).

The regressions confirm our previous findings. The choice for oneself cannot explain the choice for others and vice versa. Also, and as already discussed, choices in ME are uncorrelated across rounds whereas choices in OTHER are positively and very significantly correlated across rounds. Finally, neither age (whether computed in months or as dummy categories), gender, nor siblings have a significant effect on choices for oneself or choices for others in the second round or in all rounds together. In Appendix A4, we extend the analysis and found no effect of session size or peer composition on the choice for OTHER.

3.5 Summary

The behavior of our school-age participants is very similar across ages, except for some small differences in middle school. We have different indications of lying in our school

[20] One could argue that choices for ME and OTHER are not completely independent. Indeed, a subject could potentially under-report for others to "compensate" for one's over-report. There is no evidence of this intriguing possibility given the lack of correlation reported in section 3.2.

	ME (2)	ME (2)	ME (2)	ME (all)	OTHER (2)	OTHER (2)	OTHER (2)	OTHER (all)
ME (1)	0.071	0.067	0.064	—	0.014	0.022	0.023	—
	(0.054)	(0.055)	(0.055)		(0.055)	(0.056)	(0.056)	
OTHER (1)	0.010	0.010	0.012	—	0.169**	0.169**	0.166**	—
	(0.051)	(0.052)	(0.052)		(0.053)	(0.053)	(0.053)	
Age	—	0.001	—	—	—	-0.002	—	—
		(0.002)				(0.002)		
G2	—	—	0.260	0.439	—	—	0.074	0.003
			(0.250)	(0.360)			(0.255)	(0.395)
G3	—	—	0.142	0.489	—	—	-0.210	-0.579
			(0.236)	(0.340)			(0.242)	(0.372)
G4	—	—	0.086	0.379	—	—	-0.080	-0.090
			(0.244)	(0.352)			(0.250)	(0.386)
Male	—	0.051	0.060	0.390	—	-0.183	-0.171	-0.019
		(0.178)	(0.179)	(0.256)		(0.182))	(0.182)	(0.282)
Siblings	—	0.088	0.097	0.586	—	0.017	0.022	0.108
		(0.215)	(0.216)	(0.309)		(0.220)	(0.220)	(0.339)
Constant	4.085***	3.972***	3.899***	7.669***	2.880***	3.137***	2.970***	7.051***
	(0.293)	(0.421)	(0.352)	(0.350)	(0.300)	(0.430)	(0.360)	(0.384)
# obs.	351	351	351	351	351	351	351	351
Mult. R^2	0.005	0.006	0.009	0.024	0.029	0.034	0.036	0.009

(standard errors in parenthesis); * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 6: OLS regressions of ME and OTHER choices in second round

population. Most results are in accordance with theoretical predictions. The behavior in each round of ME provides very strong evidence of self-serving lying (for the subject's own benefit) while the strong positive correlation of OTHER choices across rounds subtly reveals the coexistence of small but statistically significant levels of altruistic lying (for the benefit of others) and spiteful lying (to the detriment of others). At the same time, some findings are surprising: there is no correlation between ME and OTHER choices and no correlation of ME choices across rounds. While these results can be simply due to a lack of power given the reduced number of rounds, they suggest that some individuals adapt to the environment to mask or exploit sequential opportunities. More generally, in light of these findings, we believe that further theoretical models and experimental data are needed to better understand the joint dynamics of lying for oneself and lying for others.

4 Lying in Middle School

A natural question at this stage is to determine the effect of the socioeconomic and intellectual background on lying. We had the opportunity to conduct the same experiment on a different population of middle schoolers. Middle school is an important transitional period from childhood to young adulthood. From a physiological viewpoint, adolescence is a critical time for the development of the brain's neural network. Physiological changes occurring during middle-school are known to impact critically cognitive and emotional responses (Choudhury et al., 2006). From an educational viewpoint, students move from small classrooms with few teachers and a close supervision to large classrooms, multiple teachers and an expectation of maturity and ability to be self-organized. From an intellectual viewpoint, the existing research using indirect (Harbaugh et al., 2001) as well as direct (Brocas et al., 2019) tests of transitivity has shown that by that age –but not earlier– children have achieved a level of basic rationality comparable to that of adults. Finally, an added benefit of studying differences in this age-group is that, as it turned out, middle schoolers at LILA behaved slightly differently from their younger and older peers.

We conducted the same experiment with a sample of 211 middle school students from Thomas Starr King Middle School (KING), a large public school located less than a mile apart from LILA. The two schools differ significantly in ethnic background (predominantly caucasian at LILA v. predominantly Latino at KING), as well school characteristics: curriculum taught (bilingual in LILA v. monolingual in KING), class size (around 20 students per class at LILA compared to 35 at KING, except in special education classes), school size (around 200 middle schoolers at LILA and 2000 at KING) and peer group (a large fraction of students at LILA remain together from pre-K to 12th grade whereas KING comprises only middle schoolers who come from many elementary schools in the Los Angeles area). For our purposes, the difference in peer group is particularly important, as LILA students are likely to know better the possible recipients of their OTHER choices than KING students.[21]

Differentiation is a core component of US public education and KING offers different tracks. For the analysis below, we follow the school (public) classification system, and separate students in "challenged" classes (110 participants) from students in "regular" classes (101 participants). The *Challenged* track is a mix of children with mild learning disabilities (dyslexia, problems focusing, etc.) and special needs (english learners), although the

[21] As Chen et al. (2016) demonstrate, even in an anonymous game, closeness to peers is likely to influence social choices.

majority are children at academic risk (low attendance and low GPA). These students are predominantly from low SES households (75% live at or below the national poverty level). The *Regular* track comprises mostly standard classes but also some honors classes (children with a GPA higher than their peers). These children come from substantially higher SES households (0% live at or below the national poverty level). Finally but importantly, we employ the *exact* same protocol as with *LILA* middle schoolers, including classroom layout, interface, instructions, payment method and conversion rate. In the analysis below, we consider the two tracks at KING separately. We include for comparison 6^{th}, 7^{th} and 8^{th} graders from the previous sample (LILA **G3**).

The study thus provides a unique chance to compare the lying propensity of 11 to 14 years old individuals as a function of socioeconomic background, demographic background and academic traits. Notice that while *Challenged* and *Regular* at KING are comparable, there is still a possible confound between academic achievement and socioeconomic status. The comparison between KING and LILA students is more complicated. Generally speaking, *LILA* exhibit academic achievements similar to *Regular* and come from households of higher SES. However, the major differences between these two groups are the school characteristics and size of peer group.

4.1 Behavior across choices

Using the same methodology as before, we first present in Table 7 some descriptive statistics of behavior in KING by track (*Challenged* and *Regular*), and add the information of LILA **G3** (*LILA*) for comparison.

	KING			LILA **G3**
	Challenged	*Regular*	*All*	*LILA*
ME-1	4.45 (1.55)	4.08 (1.81)	4.27 (1.69)	4.35 (1.45)
ME-2	4.41 (1.62)	4.23 (1.61)	4.32 (1.61)	4.47 (1.62)
ME-all	8.86 (2.53)	8.31 (2.47)	8.59 (2.51)	8.81 (2.20)
OTHER-1	3.43 (1.83)	2.92 (1.62)	3.18 (1.79)	3.22 (1.62)
OTHER-2	3.50 (1.54)	3.37 (1.65)	3.44 (1.59)	3.33 (1.61)
OTHER-all	6.93 (2.53)	6.29 (2.14)	6.62 (2.37)	6.55 (2.50)

(standard deviations in parenthesis)

Table 7: Average choices in KING by track

While the general patterns in KING are similar to those obtained in LILA, we can already notice some differences across tracks. In particular, *Regular* participants engage in less self-serving lying than *Challenged* participants (Wilcoxon rank-sum test of equality of distributions, p = 0.044). *Regular* participants also engage in spiteful lying (Kolmogorov-Smirnov test of comparison between empirical and theoretical distributions, p < 0.001) whereas their *Challenged* peers do not (Kolmogorov-Smirnov test of comparison between empirical and theoretical distributions, p = 0.066).

We next report in Figure 3 the analogue of Figure 2 to the two populations of middle schoolers at KING. Again, we include LILA middle schoolers for comparison.

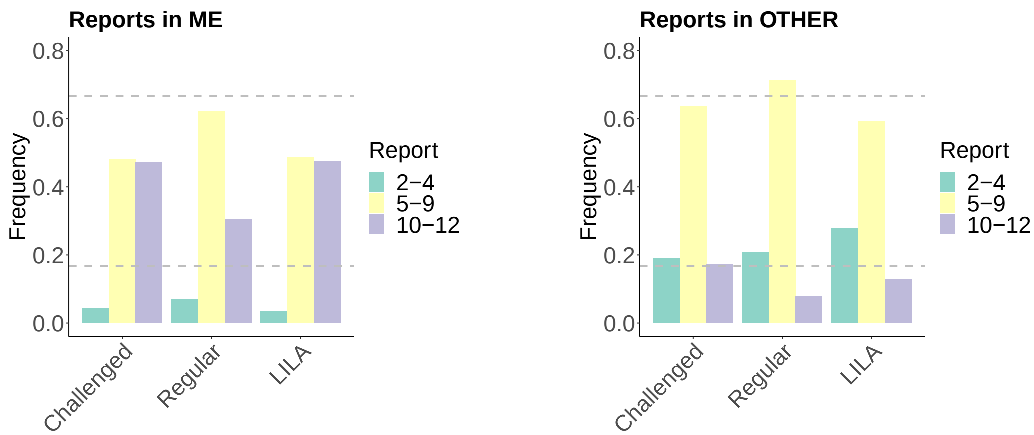

Figure 3: Allocations to ME (left) and to OTHER (right) chosen by middle schoolers

As we can see from Figure 3, looking at the distribution of behavior confirms our findings on average choices. Both *Challenged* and *Regular* lie significantly in ME (χ^2-test comparison with theoretical probabilities, p < 0.001). *Regular* lie significantly less than both *Challenged* (pairwise χ^2-test, p = 0.003) and *LILA* (pairwise χ^2-test, p = 0.001), and *Challenged* lie as much as *LILA* (pairwise χ^2-test, p = 0.833). We observe similar patterns in OTHER: *Regular* behave differently than both *Challenged* (pairwise χ^2-test, p = 0.045) and *LILA* (pairwise χ^2-test, p = 0.046), whereas *Challenged* are not different from *LILA* (pairwise χ^2-test, p = 0.077). Overall, *Challenged* participants engage more in self-serving lies and do not lie for or against others whereas *Regular* participants lie less for themselves and lie spitefully for others. In comparison, *LILA* subjects behave remarkably similar to *Challenged* in ME. Although they look between the two KING groups in OTHER, the distribution is not statistically different from that of *Challenged*.

We then report in Table 8 the analogue of Table 3a to the middle school populations.

ME \ OTHER	2-4	5-9	10-12	
	0.191	0.636	0.173	
10-12	0.127	0.254	0.091	0.478
5-9	0.064	0.354	0.064	0.478
2-4	0.000	0.027	0.018	0.045

(a) *Challenged*

ME \ OTHER	2-4	5-9	10-12	
	0.208	0.713	0.078	
10-12	0.059	0.198	0.049	0.306
5-9	0.129	0.465	0.029	0.623
2-4	0.020	0.050	0.000	0.070

(b) *Regular*

ME \ OTHER	2-4	5-9	10-12	
	0.280	0.592	0.128	
10-12	0.140	0.279	0.058	0.477
5-9	0.128	0.290	0.070	0.488
2-4	0.012	0.023	0.000	0.035

(c) *LILA*

Table 8: Reports across choices (ME and OTHER) for middle schoolers.

The result that *Challenged* subjects lie more for themselves and less against others than *Regular* subjects is confirmed in Table 8: participants in *Challenged* provide more 10-12 reports in ME than participants in *Regular* (47.8% vs. 30.6%, test of comparison of proportions, p = 0.020) and also marginally more 10-12 reports in OTHER, although the difference is not statistically significant, maybe due to a lack of power (17.3% vs. 7.8%, p = 0.067).

4.2 Behavior across rounds

We next study for each track and choice the relationship between the behavior in the first and second round. Table 9 presents the analogue of Table 4 for the population of middle schoolers (reports in first and second round of ME choices).

	L_2^m	M_2^m	H_2^m	
	0.163	0.264	0.572	
H_1^m	0.073	0.118	0.354	0.545
M_1^m	0.055	0.109	0.164	0.327
L_1^m	0.036	0.036	0.055	0.127

(a) *Challenged*

	L_2^m	M_2^m	H_2^m	
	0.208	0.307	0.485	
H_1^m	0.109	0.139	0.248	0.495
M_1^m	0.050	0.099	0.099	0.248
L_1^m	0.050	0.069	0.139	0.257

(b) *Regular*

	L_2^m	M_2^m	H_2^m	
	0.151	0.221	0.628	
H_1^m	0.058	0.128	0.314	0.500
M_1^m	0.070	0.081	0.221	0.372
L_1^m	0.023	0.012	0.093	0.128

(c) *LILA*

Table 9: Reports in ME choice across rounds for middle schoolers.

Contrary to the aggregate LILA population studied in section 3 (and to the *LILA* subgroup of middle schoolers shown here), neither group of KING subjects lie more in the second round than in the first round ($\Pr(H_1^m) = 0.545$ and $\Pr(H_2^m) = 0.572$ in *Challenged* and $\Pr(H_1^m) = 0.495$ and $\Pr(H_2^m) = 0.485$ in *Regular*). Interestingly, reports across rounds are significantly correlated for *Challenged* subjects (PCC = 0.270, p = 0.004). As discussed previously, this is what one would expect in a world where subjects have an intrinsic cost of lying. However, it is not what we found in LILA (see section 3) nor in the *Regular* population (PCC = 0.038, p = 0.704).

We can also perform the same analysis across rounds for the OTHER choice. The results are compiled in Table 10.

(a) *Challenged*

	L_2^o	M_2^o	H_2^o	
	0.290	0.427	0.282	
H_1^o	0.064	0.164	0.109	0.337
M_1^o	0.082	0.155	0.045	0.282
L_1^o	0.145	0.109	0.127	0.381

(b) *Regular*

	L_2^o	M_2^o	H_2^o	
	0.376	0.327	0.297	
H_1^o	0.099	0.050	0.050	0.198
M_1^o	0.119	0.099	0.089	0.307
L_1^o	0.158	0.178	0.158	0.495

(c) *LILA*

	L_2^o	M_2^o	H_2^o	
	0.407	0.314	0.279	
H_1^o	0.105	0.058	0.081	0.244
M_1^o	0.093	0.093	0.163	0.349
L_1^o	0.209	0.163	0.035	0.407

Table 10: Reports in OTHER choice across rounds for middle schoolers.

The results from the *Challenged* population reflect no evidence of lying for OTHERS. Aggregate proportions are close to theory (Figure 3 (right)), similar across rounds and uncorrelated (PCC = 0.126, p = 0.191). The *Regular* population exhibits spiteful lying ($\Pr(L_1^o, L_2^o) = 0.158$ v. $\Pr(H_1^o, H_2^o) = 0.050$, test for equality of proportions, p = 0.021) but there are no significant differences between rounds and no evidence of correlation across rounds (PCC = −0.142, p = 0.156). Finally, the *LILA* population also exhibits spiteful lying ($\Pr(L_1^o, L_2^o) = 0.209$ v. $\Pr(H_1^o, H_2^o) = 0.081$, test for equality of proportions, p = 0.030). We observe a positive but statistically not significant correlation across rounds (PCC = 0.198, p = 0.068).

Overall, our population of middle schoolers is rich in behavior. Not only our LILA **G3** subjects behave slightly differently from other age-groups in their school, but they also differ from participants of the same age in different tracks of another school. This age-group has a large tendency to over-report for themselves (like the other groups) as well as a small tendency to under-report for others (unlike the other groups). It is also interesting that *LILA* participants behave closer to *Challenged* than to *Regular* even though, they share more characteristics with the latter than with the former.

5 Conclusion

Our data shows that participants of all ages are willing to lie to benefit themselves but they are (mostly) reluctant to lie to benefit or hurt others. Early adolescents are slightly more prone both to self-serving and spiteful lies. These findings point to some further reflections.

Lying is a complex behavior often associated with sophisticated emotions such as guilt, shame and self-image (Battigalli et al., 2013; Utikal and Fischbacher, 2013). Recent neuroscientific research has also shown that participants who engage in lying have increased brain activity in the "control network" (Greene and Paxton, 2009), a set of prefrontal regions involved in executive control and attention which is not fully developed until adolescence and beyond (Ruff and Rothbart, 2001). The existence of only small behavioral differences across ages in our experiment indicates that young participants and adults process these emotions similarly. It suggests that lying behavior in the context of our simple paradigm does not require complex emotional trade-offs that only a fully developed control network can implement. Said differently, we are equipped from an early age with the ability to tell simple lies and process emotions around lying.

The observed differences in reports across choices derive naturally from differences in the marginal benefits of lying for oneself vs. lying for others. By contrast, the lack of correlation across those choices suggests that the cost of lying for oneself and the cost of lying for someone else are, to a large extent, independent. While this result deserves further empirical scrutiny, if the finding is general and robust, it would be an important element to guide future theoretical models of social lying. More generally, studying experimentally the dynamics of lying in a repeated choice environment could help us determine which class of models is more adequate. In particular, it could help disentangle between a standard approach where individuals are endowed with the same personal cost of lying at every period (as, for example, in Lui (1986)) and the more recent accumulation model where past lies affect the cost of current lies (as, for example, in Brocas et al. (2021)).

Finally, we have also shown small differences in the propensity to lie across populations of the same age (in our case, middle school). The most notable result is the tendency of *Challenged* students to provide higher reports in both choices than *Regular* students, resulting in more self-serving lying and no spiteful lying. This is consistent with Piff et al. (2010, 2012) who found a negative relationship between socio-economic status and prosociality, although this view has been recently challenged by Kosse et al. (2020). Surprisingly, *LILA* students behave like *Challenged* despite sharing more academic and SES characteristics with *Regular*. More generally, these differences in the middle school population should not be over-stressed as they are small in magnitude and significance, and they could simply reflect differences across populations in their marginal utility for money.

References

Johannes Abeler, Daniele Nosenzo, and Collin Raymond. Preferences for truth-telling. *Econometrica*, 87(4):1115–1153, 2019.

Sule Alan, Seda Ertac, and Mert Gumren. Cheating and incentives in a performance context: Evidence from a field experiment on children. *Journal of Economic Behavior & Organization*, 2019.

Clara Amato, Francesca Gino, Natalia Montinari, and Pierluigi Sacco. Cheating, Inequality Aversion, and Appealing to Social Norms. *Journal of Economic Behavior & Organization*, 2019.

Pierpaolo Battigalli, Gary Charness, and Martin Dufwenberg. Deception: The role of guilt. *Journal of Economic Behavior & Organization*, 93:227–232, 2013.

Laura Biziou-van Pol, Jana Haenen, Arianna Novaro, Andrés Occhipinti Liberman, and Valerio Capraro. Does telling white lies signal pro-social preferences? *Judgment and Decision Making*, 10:538–548, 2015.

Isabelle Brocas and Juan D Carrillo. A neuroeconomic theory of (dis)honesty. *Journal of Economic Psychology*, 71:4–12, 2019a.

Isabelle Brocas and Juan D Carrillo. Steps of reasoning in children and adolescents. USC Working Paper, 2019b.

Isabelle Brocas and Juan D Carrillo. Introduction to special issue "understanding cognition and decision making by children". studying decision-making in children: Challenges and opportunities. *Journal of Economic Behavior & Organization*, 179:777–783, 2020.

Isabelle Brocas, Juan D Carrillo, T Dalton Combs, and Niree Kodaverdian. The development of consistent decision-making across economic domains. *Games and Economic Behavior*, 116:217–240, 2019.

Isabelle Brocas, Juan D Carrillo, and Mallory Montgomery. Shaming as an incentive mechanism against stealing: behavioral and physiological evidence. *Journal of Public Economics*, 194(104351), 2021.

Alessandro Bucciol and Marco Piovesan. Luck or cheating? a field experiment on honesty with children. *Journal of Economic Psychology*, 32(1):73–78, 2011.

C Bram Cadsby, Ninghua Du, and Fei Song. In-group favoritism and moral decision-making. *Journal of Economic Behavior & Organization*, 128:59–71, 2016.

Jingnan Chen, Daniel Houser, Natalia Montinari, and Marco Piovesan. Beware of popular kids bearing gifts: A framed field experiment. *Journal of Economic Behavior & Organization*, 132:104–120, 2016.

Suparna Choudhury, Sarah-Jayne Blakemore, and Tony Charman. Social cognitive development during adolescence. *Social cognitive and affective neuroscience*, 1(3):165–174, 2006.

Ramon Cobo-Reyes, Jose J Dominguez, Fernando García-Quero, Brit Grosskopf, Juan A Lacomba, Francisco Lagos, Tracy Xiao Liu, and Graeme Pearce. The development of social preferences. *Journal of Economic Behavior & Organization*, 2019.

Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013.

David J Cooper and John Kagel. Other-regarding preferences. *The handbook of experimental economics*, 2:217, 2016.

Martin Dufwenberg and Martin A Dufwenberg. Lies in disguise–a theoretical analysis of cheating. *Journal of Economic Theory*, 175:248–264, 2018.

Victoria W Dykstra, Teena Willoughby, and Angela D Evans. A longitudinal examination of the relation between lie-telling, secrecy, parent–child relationship quality, and depressive symptoms in late-childhood and adolescence. *Journal of youth and adolescence*, 49 (2):438–448, 2020.

Sanjiv Erat and Uri Gneezy. White lies. *Management Science*, 58(4):723–733, 2012.

Ernst Fehr, Helen Bernhard, and Bettina Rockenbach. Egalatarianism in young children. *Nature*, 454(28):1079–1084, 2008.

Ernst Fehr, Daniela Glätzle-Rützler, and Matthias Sutter. The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. *European Economic Review*, 64:369–383, 2013.

Urs Fischbacher and Franziska Föllmi-Heusi. Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547, 2013.

Jean Gervais, Richard E Tremblay, and Diane Héroux. Boys' lying and social adjustment in pre-adolescence: teachers', peers' and self-reports. *Criminal Behaviour and Mental Health*, 8(2):127–138, 1998.

Jean Gervais, Richard E Tremblay, Lyse Desmarais-Gervais, and Frank Vitaro. Children's persistent lying, gender differences, and disruptive behaviours: A longitudinal perspective. *International Journal of Behavioral Development*, 24(2):213–221, 2000.

Matthew Gingo, Alona D Roded, and Elliot Turiel. Authority, autonomy, and deception: Evaluating the legitimacy of parental authority and adolescent deceit. *Journal of Research on Adolescence*, 27(4):862–877, 2017.

Francesca Gino, Shahar Ayal, and Dan Ariely. Self-serving altruism? the lure of unethical actions that benefit others. *Journal of economic behavior & organization*, 93:285–292, 2013.

Daniela Glätzle-Rützler and Philipp Lergetporer. Lying and age: An experimental study. *Journal of Economic Psychology*, 46:12–25, 2015.

Uri Gneezy, Agne Kajackaite, and Joel Sobel. Lying aversion and the size of the lie. *American Economic Review*, 108(2):419–53, 2018.

Joshua D Greene and Joseph M Paxton. Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences*, 106 (30):12506–12511, 2009.

William T Harbaugh, Kate Krause, and Timothy R Berry. GARP for Kids : On the Development of Rational Choice Behavior. *The American Economic Review*, 91(5): 1539–1545, 2001.

Kiryl Khalmetski and Dirk Sliwka. Disguising lies—image concerns and partial lying in cheating games. *American Economic Journal: Microeconomics*, 11(4):79–110, 2019.

Fabian Kosse, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk. The formation of prosociality: causal evidence on the role of social environment. *Journal of Political Economy*, 128(2):000–000, 2020.

JA List, R Petrie, and A Samek. How experiments with children can inform economics. Working paper, 2018.

Francis T Lui. A dynamic model of corruption deterrence. *Journal of public Economics*, 31(2):215–236, 1986.

Valeria Maggian and Marie Claire Villeval. Social preferences and lying aversion in children. *Experimental Economics*, 19(3):663–685, 2016.

Sheila Mitchell and Peter Rosa. Boyhood behaviour problems as precursors of criminalty: A fifteen-year follow-up study. *Journal of child psychology and psychiatry*, 22(1):19–33, 1981.

Paul K Piff, Michael W Kraus, Stéphane Côté, Bonnie Hayden Cheng, and Dacher Keltner. Having less, giving more: the influence of social class on prosocial behavior. *Journal of personality and social psychology*, 99(5):771, 2010.

Paul K Piff, Daniel M Stancato, Stéphane Côté, Rodolfo Mendoza-Denton, and Dacher Keltner. Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences*, 109(11):4086–4091, 2012.

Holly Alliger Ruff and Mary Klevjord Rothbart. *Attention in early development: Themes and variations*. Oxford University Press, 2001.

Mark A Stewart and Susan de Blois. Diagnostic criteria for aggressive conduct disorder. *Psychopathology*, 18(1):11–17, 1985.

Magda Stouthamer-Loeber. Lying as a problem behavior in children: A review. *Clinical Psychology Review*, 6(4):267–289, 1986.

Matthias Sutter, Claudia Zoller, and Daniela Glätzle-Rützler. Economic behavior of children and adolescents–a first survey of experimental economics results. *European Economic Review*, 111:98–121, 2019.

Jean Tirole. A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *The Review of Economic Studies*, 63(1):1–22, 1996.

Verena Utikal and Urs Fischbacher. Disadvantageous lies in individual decisions. *Journal of Economic Behavior & Organization*, 85:108–111, 2013.

Shanshan Zhang, Narek Bejanyan, Zhou Fang, Matthew Gomies, Jason Justo, Hsin Li, Rainita Narender, Minjae Yun, and Joshua Tasoff. Trick for a treat: The effect of costume, identity, and peers on norm violations. CGU Working Paper, 2019.

Appendix A - Additional analysis

A1. Power Analysis

We report the power analysis conducted ex ante–before running the experiment–but after knowing the approximate size of the sample we had access to: between 70 and 110 per age group. The main hypothesis tested is whether the *distribution* of reports is consistent with truthful reports (our null hypothesis). We looked at several specifications of that same test, and report the one that corresponds to the three bins of sum of reports in rounds 1 and 2 (2-4, 5-9 and 10-12) that we retained for the analysis (see section 3.1).

(Ex ante) Power calculations

The power analysis looks for the effect size we can detect if we test $\mathbf{H_0}$ at $\alpha = 0.05$ with a sample size between 70 and 110, and if we want to achieve a power of $1 - \beta = 0.8$. In our specification with 3 bins, there are 2 degrees of freedom (df = 2). We can detect an effect size of 0.37 (or more) with $n = 70$ and 0.29 (or more) with $n = 110$. Calculations for other values of power and effect size are reported in Figure 4.

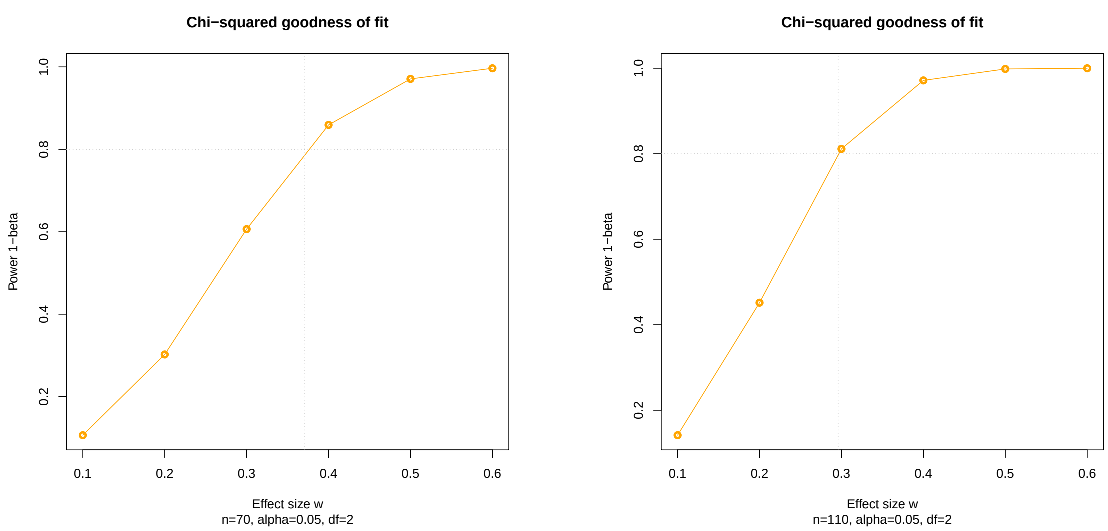

Figure 4: Power Calculations

(Ex post) Effect sizes

The effect sizes associated with the empirical frequencies displayed in Figure 2 are reported in Table 11, broken-down by age-group and choice.[22]

	G1	**G2**	**G3**	**G4**	**G-All**	**U**
ME	0.56	0.58	0.85	0.76	0.68	0.54
OTHER	0.17	0.05	0.30	0.22	0.14	0.14

Table 11: Effect sizes for ME and OTHER by age-group

According to the heuristic classification in Cohen (2013), effects are large in ME and small in OTHER, with the exception of **G3**, which is borderline. Effects are similar across school-age groups, so also similar when we pool all school-age participants together (**G-All**). The effects are in the same range (but typically smaller) in our control adult group.

Consider the special case of students in **G3** who report draws 2-4, 5-9 and 10-12 with respective frequencies 0.03, 0.49 and 0.48 in ME. The very large effect size of 0.85 corresponds to 82% and 26% fewer low and medium reports than expected and 188% more high reports than expected. In OTHER, the reported frequencies 0.28, 0.59 and 0.13 correspond to 68% more low reports than expected, 11% fewer medium reports and 22% fewer higher reports than expected. The dispersion around expected frequencies is more moderate, resulting in a lower effect size. If, instead, we consider the case of **G2** in OTHER, the reported frequencies are 0.15, 0.70 and 0.15, which are almost identical to the expected frequencies, hence an almost inexistent effect size.

A2. Disaggregated reports

Figure 5 presents the same information as Figure 2 –the total allocation to ME and OTHER by age group– except that we consider every report 2 to 12 separately, instead of grouping them in three categories.

When we compare empirical and theoretical distributions using the disaggregate data, we obtain the same conclusions as with the more aggregate analysis of Figure 2. The overall distribution is statistically different from the theoretical triangular distribution for all age groups in ME (Kolmogorov-Smirnov test of differences in distributions, $p < 0.001$). The distribution in OTHER is different from theory in **G3** ($p = 0.010$) but not in the other

[22]We use the following formula for calculating the effect size:

$$w = \sqrt{\sum_i \frac{(p_i - q_i)^2}{q_i}}$$

where q_i is the expected frequency of observations in bin i, which in our case is $(1/6, 2/3, 1/6)$, and p_i is the empirical frequency.

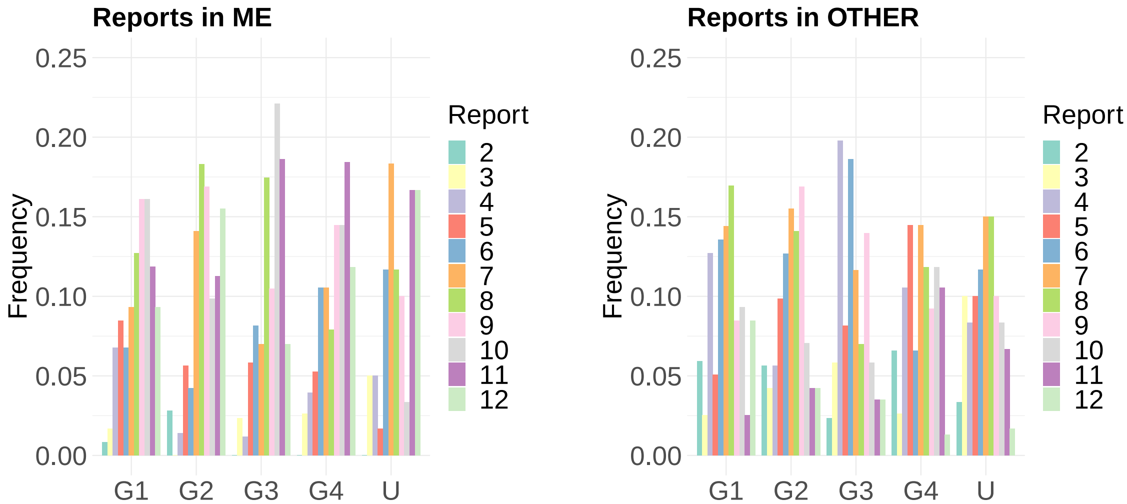

Figure 5: Disaggregate allocations by age group

age groups ($p > 0.05$). Looking at reports separately, in ME it seems that **G3** are attracted by 10 whereas **U** stick to 11 and 12. In OTHER, 4 is over-represented in **G3** whereas **U** is remarkably close to theory. However, these are only visual differences: with an average of 80 subjects per age group and a non-observed game, we do not have enough power to conduct significance tests on each report separately. In any case, the data reinforces the idea that deviations are not limited to corner reports (1 or 6).

A3. First period honest individuals

We can also study in more detail the choices of participants who reported 1 or 2 in the first round of ME (L_1^m). Indeed, it is highly unlikely that these individuals have provided dishonest reports. Among LILA children, there are 64 individuals fitting this description split as follows: 24 in **G1**, 16 in **G2**, 11 in **G3** and 13 in **G4** (28 male and 36 female). Table 12 describes the proportion of reports in the second round of ME (R_2^m) as well as in both rounds of OTHER (R_1^o and R_2^o).

ME-2 (R_2^m)			OTHER-1 (R_1^o)			OTHER-2 (R_2^o)		
L_2^m	M_2^m	H_2^m	L_1^o	M_1^o	H_1^o	L_2^o	M_2^o	H_2^o
0.187	0.264	0.549	0.359	0.328	0.313	0.344	0.328	0.328

Table 12: Behavior of participants who reported L_1^m

We performed a Kolmogorov-Smirnov test comparing the empirical and theoretical distributions. Behavior in round 2 of ME is statistically different from uniform ($p = 0.005$) whereas behavior in OTHER is not ($p = 0.809$ and $p = 0.952$ in rounds 1 and 2).

The proportion of R_2^m choices is just a restatement of the last line in Table 4b. As already discussed, and to our great surprise, these individuals exhibit similar levels of lying in the second round of ME compared to their peers. It reinforces the documented lack of correlation in ME choices across rounds. By comparing R_1^o and R_2^o with the proportions in the first line and last column of Table 5a, we notice that the behavior of these individuals is also very similar to the rest of the population in OTHER. As one would expect, if these individuals do not lie for themselves (at least in round 1), they are also unlikely to lie for others in either round.

A4. Peer effects

Given school constraints, we had to adjust our sessions to accommodate teacher requests. This resulted in some minor differences across sessions. First, sessions differed in the peer group: 22 sessions had all participants from the same grade while 8 sessions had a mix of participants from two consecutive grades.[23] Second, sessions differed in size: 26 sessions had 12 participants, 3 sessions had 9 participants and 1 session had 15 participants (remember that we had a total of 354 subjects but we lost the data for 3 of them).

These differences are unlike to affect ME choices but they could have an influence on OTHER choices. Indeed, size may compromise anonymity and familiarity with other participants may affect the behavior towards them. To study this possibility we ran the same regressions on OTHER choices as in columns 7 and 8 of Table 6, except that we added dummy variables for peer group (*MixedGrade* = 1, for sessions with a mix of participants from consecutive grades) and session size (*Session9* and *Session15*, with *Session12* being the omitted category). Results are reported in Table 13.

Mixing grades and having sessions of different size had no significant effect on OTHER choices. This is not overly surprising. Indeed, LILA is a small school with 40 to 60 students per grade. Some children stay together from elementary to high school, so they are familiar with everyone in the school and develop strong friendship ties with people outside their grades. More generally, no student in the experiment felt isolated in a foreign environment. As for session size, changes were minor, since the majority of sessions had exactly 12 subjects. The absence of a statistical effect is, therefore, also expected.

[23] These sessions were either evenly split, or with a slight majority from one grade over the other. We never had only one or two participants from a different grade alone in a session.

	OTHER (2)	OTHER (2)	OTHER (all)	OTHER (all)
ME (1)	0.024	0.020	—	—
	(0.362)	(0.057)		
OTHER (1)	0.169**	0.165**	—	—
	(0.053)	(0.054)		
G2	0.089	0.097	-0.022	0.048
	(0.256)	(0.258)	(0.397)	(0.398)
G3	-0.211	-0.232	-0.576	-0.702
	(0.242)	(0.247)	(0.373)	(0.379)
G4	-0.059	-0.151	-0.127	-0.225
	(0.252)	(0.271)	(0.390)	(0.418)
Male	-0.184	-0.180	0.001	-0.001
	(0.184)	(0.184)	(0.283)	(0.282)
Siblings	0.013	0.008	0.122	0.131
	(0.221)	(0.221)	(0.340)	(0.340)
MixedGrade	-0.144	-0.108	0.240	0.260
	(0.210)	(0.214)	(0.324)	(0.329)
Session9	—	0.163	—	0.948
		(0.361)		(0.555)
Session15	—	0.586	—	0.977
		(0.527)		(0.812)
Constant	2.998***	3.012***	6.980***	6.897***
	(0.362)	(0.363)	(0.397)	(0.398)
# obs.	351	351	351	351
Mult. R^2	0.014	0.041	0.010	0.023

(standard errors in parenthesis); * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 13: Extended OLS regressions of OTHER choices in second round

Appendix B - Instructions

[Instructions for participants in middle school and above. Instructions in elementary school (kindergarten to 5th grade) are the same except for payment method]

Hi, we are going to play a few games. In all the games, you will earn points that will be placed in your virtual wallet. At the end of the experiment you will be paid 1 cent for each point you get. You will receive your payment with an Amazon gift card sent to your school email address. You will get several hundred points, so you will be able to get a nice gift card. In all the games, you will play through the tablets. We ask you to not talk and keep your decisions private.

Dice Game

We will first play a "dice game" to determine how much you get just for being here today. You have a red dice and a green dice in the cup we gave you. You will shake the cup and overturn it on your table. You will then lift the cup to reveal the numbers. Make sure to keep the numbers secret. Now, the number on the red dice corresponds to points for yourself while the number on the green dice corresponds to points for another person in the room. You have to input the numbers on the computer. The screen will show this.

[SLIDE 1]

To input a number, you just need to tap on the corresponding dice and press OK. The computer will record the numbers and multiply them by 30. This screen shows how many points are earned for each number clicked and how much money this represents.

[SLIDE 2]

Here is something important about the game. You do not know and will not know who the other person is. The computer will choose. You just know that the person who receives these points is in this room and is not the person you receive points from. So, this is not an exchange.

Look at the example here.

[SLIDE 3]

Each student receives points from one person and gives points to a different person.

Go ahead, shake the cup and reveal the secret numbers. Answer on the computer and press OK. Once you press OK, the screen will disappear.

The computer will now select to which student your points go, record the points and place them in your wallet.

Remember that these are the points you got just for being here. Now, we are going to play the exact same game one more time. This will be the last time we play this game.

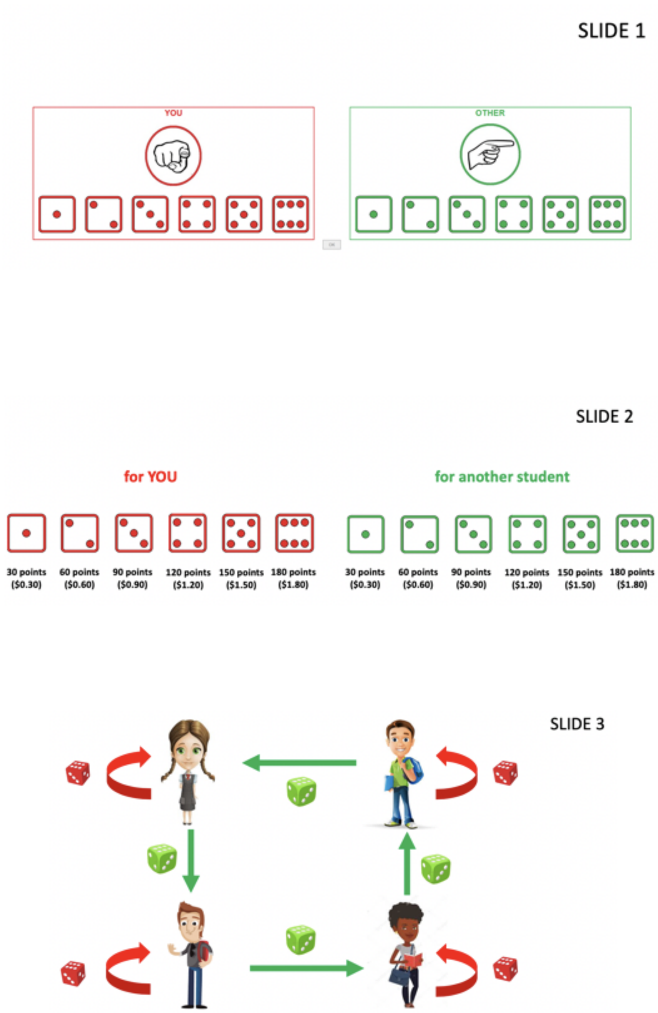

Figure 6: Slides projected during instructions.